PATENT ABSTRACTS OF JAPAN

(11)Publication number : 10-098655
(43)Date of publication of application : 14.04.1998

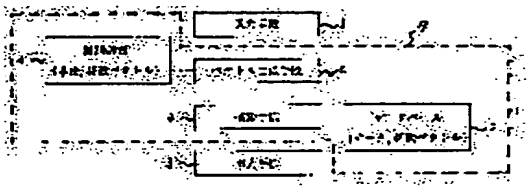
(51)Int.Cl. H04N 5/44
G11B 15/02
H04N 5/445

(21)Application number : 08-252119
(22)Date of filing : 25.09.1996

(71)Applicant : SHARP CORP
(72)Inventor : KUROMUSHIYA KENICHI
IKEUCHI HIROSHI
KARASHI IKUO

(54) PROGRAM RETRIEVAL DEVICE

(57)Abstract:
PROBLEM TO BE SOLVED: To provide a program retrieval device with a small calculation amount for the retrieval by which programs such as TV, CTV, teletext and radio programs are retrieved and reservation of programs are made simply and the retrieval is conducted even when a retrieval keyword is not completely accurate.
SOLUTION: The device is provided with a database 2 in which program data such as program date and time, channel, program name, contents of program and performers are stored. A word relating to a desired program is entered by an entry means 1, a word dictionary 4 having a characteristic vector of each word is referenced, a characteristic vector corresponding to a question text from the entry means 1 is generated by a vector generating means 5, a retrieval means 6 calculates a distance between the characteristic vector corresponding to the program data in the database with the generated characteristic vector so as to retrieve the desired program.



LEGAL STATUS

[Date of request for examination] 14.02.2000
[Date of sending the examiner's decision of rejection]
[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]
[Date of final disposal for application]
[Patent number]
[Date of registration]
[Number of appeal against examiner's decision of rejection]
[Dat of requesting appeal against examiner's decision f rejection]
[Date of extinction of right]

[Title of the Invention] Program Searching Apparatus

[Summary]

[Object] When searching a program or programing a video recording, a lot of time and labor are required, which causes input mistakes and long searching time.

[Solving Means] A program searching apparatus is provided with a database 2 in which program data such as date and time of programs, channels, program titles, program contents and performers is stored and a user inputs a word related to a desired program from inputting means 1, refers to a word dictionary 4 having a characteristic vector of each word, prepares a characteristic vector corresponding to a question text from the inputting means 1 by vector generating means 5 and calculates a distance between the characteristic vector and a characteristic vector corresponding to program data in the above-mentioned database to search the desired program by searching means 6.

[0013]

[Embodiment Mode of the Invention] A program searching apparatus of the present invention utilizes a searching device that uses a characteristic vector. A configuration of the program searching apparatus using a characteristic vector is shown in Fig. 1. This program searching apparatus is composed of inputting

means 1, which consists of a keyboard, a tablet and a microphone with which a user inputs a question text including contents semantically close to contents of a desired program, a database 2 in which program data and a characteristic vector corresponding to each piece of the program data are stored in pairs, outputting means 3 consisting of a CRT and a printer for outputting a search result, a word dictionary 4 in which words and a characteristic vector corresponding to each word are stored in pairs, vector generating means 5 for generating a vector based on a word inputted from the inputting means 1 and the word dictionary 4 and searching means 6 for executing predetermined calculation based on a result of vector generation by the vector generating means 5 and the database 2. A searching device 7 is composed of a general-purpose computer, a memory and an external storage device.

[0014] First, the characteristic vector will be described. The characteristic vector is a vector indicating a relation between a concept that a word in a sentence has and a context, and represents a degree of semantic relation with a multiplicity of feature words as a vector. For the characteristic vector of the present invention, a text searching technology by a context vector is used, which is described in "Association search from large-scale database", Shigakugiho AI92-99 (1993-1) issued by the Institute of Electronics, Information and Communication Engineers. That is, the "characteristic vector" in this

embodiment mode directly corresponds to the above-mentioned "context vector".

[0015] Assuming that k concept classifications are feature words, a value of each element of a k-dimensional vector is associated with respective feature words. A value of each element of a context vector Xi = (xi1, xi2, ..., xik) of a word i is $0 \leq xij \leq Em$. Em is a positive constant. If there is no relation between the word i and a feature word j, xij=0, and if there is a relation between them, xij takes a large value according to a degree of the relation. For example, assuming that a characteristic vector consists of five feature words (nature, city, noise, animal, green), if a value of each element is one of two values, 0 and 1, a characteristic vector of a word "mountain" can be expressed as (1, 0, 0, 1, 1) or the like.

[0016] Next, procedures of a program search will be described. A user inputs a question text having contents semantically close to contents of a desired program from the inputting means 1. The question text may be inputted word by word or inputted using a sentence or a natural language as long as it contains words. As shown in Fig. 7, if an inputted question text is a sentence or the like, the vector generating means 5 extracts respective words, reads out a characteristic vector corresponding to each word from the word dictionary 4 and calculates a sum of those vectors to normalize each vector. The searching means 6 calculates a

distance between the sum of the vectors and a characteristic vector corresponding to each piece of program data in the database 2 and outputs the program data to the outputting means 3 as a search result in order from a piece of the program data having a characteristic vector closest in distance to the sum.

[0017] A distance calculation in the searching means 6 will be described more specifically. In the present invention, an inner product is calculated for calculating a distance. If two vectors are X=(x1, x2, ..., xm) and Y=(y1, y2, ..., ym), inner product X·Y is expressed as X·Y = x1 x y1 + x2 x y2 + ... + xm x ym. This means that the larger this inner product value the closer the distance. For example, assuming that a characteristic vector Q of a question text q is expressed as Q = (3, 5, 4, 2, 4, 5, 2, 1) and characteristic vectors S and T of data s and t in the database 2 are expressed as:

S = (4, 5, 4, 1, 4, 5, 0, 1)

T = (5, 0, 4, 6, 3, 1, 3, 2),

Q·S = 3 x 4 + 5 x 5 + 4 x 4 + 2 x 1 + 4 x 4 + 5 x 5 + 2 x 0 + 1 x 1 = 97

Q·T = 3 x 5 + 5 x 0 + 4 x 4 + 2 x 6 + 4 x 3 + 5 x 1 + 2 x 3 + 1 x 2 = 68, which means that the question text q is more closer in distance to the data s than to the data t.

[0018]

[Embodiment] A program searching apparatus in accordance with an

embodiment of the present invention is shown in Fig. 2. Here, parts identical with those in Fig. 1 are denoted by the identical reference numerals. Compared with Fig. 1, data inputting means 8 for inputting data in the database 2, wired or wireless communicating means 9 for controlling an apparatus to be controlled 10 and sending program data and an apparatus to be controlled 10 are added to the searching device 7. The apparatus to be controlled 10 is a receiver such as a television, a CATV receiver, a teletext broadcasting receiver and a radio, or a videocassette recorder. The searching device 7 outputs all or a part of program data, which is selected by the inputting means 1 out of candidates of program data in a search result outputted to the outputting means 3, to the apparatus to be controlled 10 via the communicating means 9.

[0019] Program data such as a date, a starting time, an ending time, a channel, a program title, program contents and performers for each program is inputted in the searching device 7 from the data inputting means 8 and stored in the database 2. An example of a data configuration of these pieces of program data is shown in Fig. 3.

[0020] Next, preparation of a program database of the database 2 will be described. An input from the data inputting means 8 is performed by means of an input from a keyboard, an OCR input, a voice input, an input from online data such as personal computer

communication, teletext broadcasting data receipt and the like. Inputted program data is converted to a characteristic vector for each piece of program data and stored in the database 2 as a pair of the program data and the characteristic vector. In this context, respective words are extracted from the program data by the function of the vector generating means 5, a characteristic vector corresponding to each word is read out from the word dictionary 4, and a sum of the vectors is calculated and normalized such that sizes of the vectors are constant, whereby the characteristic vector is generated.

[0021] Processing for replacing this program data with a characteristic vector will be described more specifically. This processing is the same as the processing for generating a characteristic vector from words. Here, the case in which contents of program data is as shown in Fig. 4 and a part of contents of the word dictionary 4 is as shown in Fig. 5 will be considered. Here, a program title, program contents and other parts of the data are extracted from the program data and particles are removed to extract each word. In this context, "news", "today", "special story", "spring" and "topics" are extracted. Next, characteristic vectors for these words are read out from the word dictionary 4 to calculate a sum Vs of these vectors. If a characteristic vector of a word X is represented as V(X), Vs = V (news) x 2 + V (today) + V (special story) + V (spring) + V (topics).

Since "news" is extracted in two parts, it is multiplied by two. When values of the characteristic vectors in Fig. 5 are substituted in this expression, Vs = (3, 5, 0, 4, 2, ..., 3). Next, this vector is normalized. If the normalized vector is Vn, Vn = a x Vs/|Vs|. Here, a is a size of a vector after it is normalized and |Vs| is a size of the vector Vs and takes a positive value and, if Vs = (v1, v2, ..., vm), $|Vs|^2 = |v1|^2 + |v2|^2 + \ldots + |vm|^2$. Here, the normalization is performed on the assumption that a = 10. The characteristic vectors normalized in this way and the program data are stored in the database 2 as a pair.

[0022] Fig. 6 is a flow chart of processing of a program search. First, a user inputs from the inputting means 1 a question text indicating search conditions such as program title and program contents that are semantically close to program contents that the user wishes to search (S1). Program data in the database 2 is searched by the searching device 7 (S2). Since the search performed here is an association search, program data of contents semantically close to the desired program contents can be searched even if data completely coinciding with it does not exist in the program data. A search result of the searching device 7 is outputted to the outputting means 3 (S3) to confirm and select contents of the search (S4). The contents of the search is selected by designating desired program data using the inputting means 1 out of candidates of program data displayed on, for example,

a CRT. Then, necessary program data in the program data selected out of the search result is sent to the apparatus to be controlled 10 via the communicating means 9. In case of setting and changing a channel (S5), data of a channel in the program data is sent to the apparatus to be controlled 10 and a channel of the apparatus to be controlled 10 is set and changed based on the received channel. In addition, if a program has not started yet, data of a date and a starting time, and an ending time if necessary, in the program data is sent to and set in the apparatus to be controlled 10. On the other hand, in case of programing a video recording (S5), data of a channel, a date and a starting time, and an ending time if necessary, in program data that a user wishes to reserve for video recording is sent to the apparatus to be controlled 10 and the programing for a video recording of the apparatus to be controlled 10 is set based on the received data. Since program contents can be confirmed by outputting program data to the outputting means 3 at the time of programing for the video recording, it is possible to confirm if there is any mistake in contents of the programing of the video recording.

[0023] Since the searching device 7 performs a search using a characteristic vector, it is possible to perform a semantic search to find out a program. For example, if "comedy program" is inputted from the inputting means 1, programs including data of "comedy", "comic dialogue", "comic monologue" and the like can

be searched in addition to those including "comedy program" in program data. In this way, even if contents of a target program is not specifically inputted, a program having contents semantically close to contents of the target program can be searched. Moreover, a search result can be narrowed and a target program can be found out by adding conditions such as a date, a day of the week, a time, performers and a channel.